Attorney Docket No. 66161 (7114)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appln No.:	09/692,846
Applicants:	Konopka, Courtney C.
Filed:	October 19, 2000
For:	NATURAL LANGUAGE INTERFACE CONTROL SYSTEM
TC/A.U.:	2654
Examiner:	Lamont Spooner

This document was filed electronically using the USPTO's EFS-WEB.

APPEAL BRIEF

Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Pursuant to 37 C.F.R. §1.192, the applicants hereby respectfully submit the following Brief in support of their appeal.

(1) Real Party in Interest

The real parties in interest are (a) Sony Corporation, a Japanese corporation having a primary place of business in Tokyo, Japan; and (b) Sony Electronics Inc., a U.S. corporation having a primary place of business in Park Ridge, New Jersey.

(2) Related Appeals and Interferences

No related appeals or interferences are known to the Appellant.

(3) Status of Claims

Claims 1-17, 26-30, and 32-44 are pending. All of the claims are under final rejection.

(4) Status of Amendments

No amendments have been submitted subsequent to the Final Rejection in this application.

(5) Summary of Claimed Subject Matter

In the pending application, claims 1-17, 26-30, and 32-44 are pending. Claims 18-25, 31, and 45-56 have been previously cancelled. Claims 1, 6, 7, 8, 9, 10, 17, and 26 are independent claims and the remaining claims are dependent claims.

In previous systems attempts at natural language processing of human speech have been both inefficient and rigid. For example, natural language interfaces have been implemented as automated phone systems such as those used by airline reservation systems. Such systems prompt the user to speak within a certain context. In such systems, the received speech must be in predetermined and fixed format in order that the speech can be used by the system.

These previous approaches suffer from several disadvantages. For example, since the received speech must be in a fixed format, the use of open-ended requests, that is, requests unrestricted according to a form, format, or syntax, are unsupported by these previous systems. In fact, when an open-ended request was received, previous systems typically either ignored the request or reported an error to the user making the request.

The Applicants' invention addresses the shortcomings and limitations of previous systems. More specifically, independent claim 1 recites an interface control system for operating a plurality of devices. The system includes a 3 dimensional microphone array (e.g., arrays 108 as shown in FIG. 2 of the Application, reproduced below for the convenience of the reader) and a feature extraction module (e.g., feature extraction module 202) coupled to the first microphone array (e.g., array 108). A speech recognition module (e.g., speech recognition module 204) is coupled to the feature extraction module (e.g., feature extraction

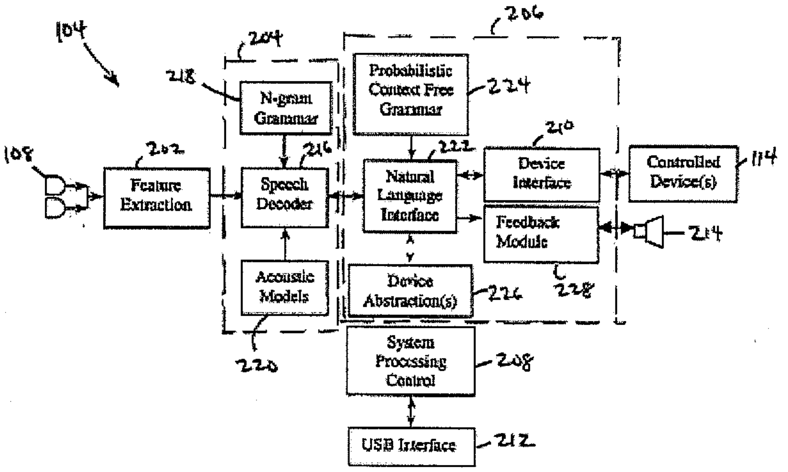

FIG. 2

module 202) and the speech recognition module utilizes hidden Markov models and can switch between different acoustic models and different grammars. Specification, page 14, lines 16- 34. In one example, these different grammars are the rules by which lexica are built with the lexica being dictionaries consisting of words and their pronunciation entries. Specification, page 19, lines 12-32. At least one of the different acoustic models and at least one of the different grammars is downloaded over a network. Specification, page 11, lines 29-34. A natural language interface module (e.g., natural language control module 206) is coupled to the speech recognition module (e.g., speech recognition module 204). A device interface (e.g., device interface 212) is coupled to the natural language interface module (e.g., natural language control module 206) and the natural language interface module operates a plurality of devices (e.g., devices 114) of one or more types that are coupled to the device interface based upon non-prompted, open-ended natural language requests from a user. Specification, page 5, line 25- page 6, line 22. The natural language interface module abstracts each of the plurality of devices into a respective one of the different grammars and a respective one of a plurality of lexica corresponding to each of the plurality of devices. Specification, page 10, line 30- page 11, line 6.

Independent claim 6 recites a natural language interface control system that operates a plurality of devices. The system includes a 3 dimensional microphone array (e.g., array 108) and a feature extraction module (e.g., feature extraction module 202) that is coupled to the microphone array (e.g., array 108). A speech recognition module (e.g., speech recognition module 204) is coupled to the feature extraction module (e.g., feature extraction module 202) and the speech recognition module (e.g., speech recognition module 204) utilizes hidden Markov models and can switch between different acoustic models and different grammars. Specification, page 14, lines 16- 34. A natural language interface module (e.g., natural language control module 206) is coupled to the speech recognition module and a device interface (e.g., device interface 212) is coupled to the natural language interface module (e.g., natural language control module 206). The natural language interface module (e.g., natural language control module 206) operates a plurality of devices of one or more types (e.g., devices 114) that are coupled to the device interface (e.g., device interface 212) based upon non-prompted, open-ended natural language requests from a user. The natural language interface (e.g., natural language control module 206) abstracts each of the plurality of devices into a respective one of a plurality of grammars and a respective one of a plurality of lexica corresponding to each of the plurality of devices. Specification, page 10, line 30- page 11, line 6.

Independent claim 7 recites a natural language interface control system for operating a plurality of devices. The system includes a 3 dimensional microphone array (e.g., array 108) and a feature extraction module (e.g., feature extraction module 202) that is coupled to the microphone array (e.g., array 108). A speech recognition module (e.g., speech recognition module 204) is coupled to the feature extraction module (e.g., feature extraction module 202) and the speech recognition module (e.g., speech recognition module 204) utilizes hidden Markov models and can switch between different acoustic models and different grammars. Specification, page 14, lines 16- 34. A natural language interface module (e.g., natural language control module 206) is coupled to the speech recognition module and a device interface (e.g., device interface 212) is coupled to the natural language interface module (e.g., natural language control module 206). The natural language interface module (e.g., natural language control module 206) operates a plurality of devices of one or more types (e.g., devices 114) that are coupled to the device interface (e.g., device interface 212) based upon non-prompted, open-ended natural language requests from a user. The natural language interface module (e.g., natural language control module 206) searches for the non-prompted, open-ended user requests upon the receipt and recognition of an attention word. Specification, page 10, lines 14-29.

Independent claim 8 recites a natural language interface control system for operating a plurality of devices. The system includes a 3 dimensional microphone array (e.g., array 108) and a feature extraction module (e.g., feature extraction module 202) that is coupled to the microphone array (e.g., array 108). A speech recognition module (e.g., speech recognition module 204) is coupled to the feature extraction module (e.g., feature extraction module 202) and the speech recognition module (e.g., speech recognition module 204) utilizes hidden Markov models and can switch between different acoustic models and different grammars. Specification, page 14, lines 16- 34. A natural language interface module (e.g., natural language control module 206) is coupled to the speech recognition module and a device interface (e.g., device interface 212) is coupled to the natural language interface module. The natural language interface module (e.g., natural language control module 206) operates a plurality of devices of one or more types (e.g., devices 114) that are coupled to the device interface (e.g., device interface 212) based upon non-prompted, open-ended natural language requests from a user. The natural language interface module (e.g., natural language control module 206) context switches grammars, acoustic models, and lexica upon receipt and recognition of an attention word. Specification, page 20, line 13- page 21, line 12.

Independent claim 9 recites a natural language interface control system for operating a plurality of devices. The system includes a 3 dimensional microphone array (e.g., array 108) and a feature extraction module (e.g., feature extraction module 202) that is coupled to the microphone array (e.g., array 108). A speech recognition module (e.g., speech recognition module 204) is coupled to the feature extraction module (e.g., feature extraction module 202) and the speech recognition module (e.g., speech recognition module 204) utilizes hidden Markov models and can switch between different acoustic models and different grammars. Specification, page 14, lines 16- 34. A natural language interface module (e.g., natural language control module 206) is coupled to the speech recognition module (e.g., speech recognition module 204) and a device interface (e.g., device interface 212) is coupled to the natural language interface module (e.g., natural language control module 206). The natural language interface module (e.g., natural language control module 206) operates a plurality of devices of one or more types (e.g., devices 114) that are coupled to the device interface (e.g., device interface 212) based upon non-prompted, open-ended natural language requests from a user. A grammar module (e.g., grammar module 218) stores different grammars for each of the plurality of devices.

Independent claim 10 recites a natural language interface control system for operating a plurality of devices. The system includes a 3 dimensional microphone array (e.g., array 108) and a feature extraction module (e.g., feature extraction module 202) coupled to the microphone array (e.g., array 108). A speech recognition module (e.g., speech recognition module 204) is coupled to the feature extraction module and the speech recognition module (e.g., speech recognition module 204) utilizes hidden Markov models and can switch between different acoustic models and different grammars. Specification, page 14, lines 16- 34. A natural language interface module is coupled to the speech recognition module (e.g., speech recognition module 204) and a device interface (e.g., device interface 212) is coupled to the natural language interface module (e.g., natural language control module 206). The natural language interface module (e.g., natural language control module 206) operates a plurality of devices of one or more types (e.g., devices 114) that are coupled to the device interface (e.g., device interface 212) based upon non-prompted, open-ended natural language requests from a user. An acoustic model module (e.g., acoustic models 220) stores different acoustic models for each of the plurality of devices (e.g., devices 114).

Independent claim 17 recites searching for an attention word based on a first context including a first set of models, grammars, and lexica. Upon finding the attention word, the first context is switched to a second context in order to search for an open-ended user request. For example the open-ended user requests may include "I wanna watch TV", "hey, let's watch TV", "Turn on the TV", or "Do you have the have the album 'Genesis'?" The second context includes a second set of models, grammars, and lexicons. See Specification, page 6, line 23- page 8, line 4.

Independent claim 26 recites a natural language interface control system for operating a plurality of devices. The system includes a first microphone (e.g., array 108) and a feature extraction module (e.g., feature extraction module 202) that is coupled to the first microphone (e.g., array 108). A speech recognition module (e.g., speech recognition module 204) is coupled to the feature extraction module (e.g., feature extraction module 202) and a natural language interface module (e.g., natural language control module 206) is coupled to the speech recognition module (e.g., speech recognition module 204). A device interface (e.g., device interface 212) is coupled to the natural language interface module (e.g., natural language control module 206), and the natural language interface module (e.g., natural language control module 206) operates a plurality of devices of one or more types (e.g., devices 114) that are coupled to the device interface (e.g., device interface 212) based upon non-prompted, open-ended natural language requests from a user. An external network interface is coupled to the natural language interface control system. The natural language interface (e.g., natural language control module 206) abstracts each of the plurality of devices into a respective one of a plurality of grammars and a respective one of a plurality of lexica corresponding to each of the plurality of devices. Specification, page 10, line 30- page 11, line 6.

(6) Grounds of Rejection to be Reviewed on Appeal

(A) Whether claim 17 is anticipated by U.S. Patent No. 6,584, 439 to Geilhufe?

(B) Whether claims 1-16, 26-30, and 32-44 are unpatentable under 35 U.S.C. §103 over U.S. Patent No. 6,324,512 to Junqua in view of an article by Giuliani ("Hands Free Continuous Speech Recognition in Noisy Environment Using a Four Microphone Array") and U.S. Patent No. 6,408,272 to White ("the White patent")?

(7) Argument

(A) Claim 17 is Not Anticipated by Geilhufe

As mentioned, claim 17 recites searching for an attention word based on a first context including a first set of models, grammars, and lexica. Upon finding the attention word, the first context is switched to a second context to search for an open-ended user request. The second context includes a second set of models, grammars, and lexica.

The Examiner stated that Geilhufe teaches the search for an open-ended user request. Specifically, the Examiner stated that the phrase "Aardvark Call Mom" (received by the Geilhufe system) was an open-ended user request. In addition, the Examiner asserted that

> the personal name of Aardvark employs its own grammar, lexicon, and model of device names- which is inherently and undeniable a context, wherein the user must supply the word Aardvark- wherein in the context is interpreted as the application determination, secondly, the context is inherently switched to a second context (or topic) directly relating to the open ended user request, this second context employs only thereafter a second grammar, model and lexicon which it accesses after the keyword "Aardvark" is determined.

The Applicants respectfully disagree with these statements for the reasons stated below.

As an initial matter, the Geilhufe system is not able to process open-ended user requests. In fact, the requests received must be in a predetermined format or the Geilhufe system will not be able to recognize them. More specifically, Geilhufe describes a standard command syntax that is used "for ***all*** voice commands." See Geilhufe, col. 19, lines 15-37 (emphasis added).[1] While Geilhufe mentions the existence of two alternative command formats, only a single format is ever used. See Geilhufe, col. 19, lines 55-67 and col. 20, lines 29-36. In other words, all commands of Geilhufe must follow a fixed format and cannot deviate from the standard format, whatever that format is.

For these reasons, the Applicants assert that Geilhufe fails to teach or suggest the searching for an open-ended user request as recited in claim 17 and, consequently, claim 17 is not anticipated by Geilhufe.

In addition, the Geilhufe system does not use or switch between a first context (having a first set of models, grammars, and lexica), and a second context (having a second set of models, grammars, and lexica) upon finding an attention word as recited in claim 17. In rejecting claim 17, the Examiner analyzed the phrase "Aardvark call Mom" and asserted that this phrase could be split into two portions. A first portion ("Aardvark") was deemed by the Examiner to be an attention word and a second portion ("call mom") was deemed to be an open-ended command. Furthermore, the Examiner stated that each of the two portions

[1] The Geilhufe system mandates that user requests must be in the form of <silence> <name> <command> <modifiers & variables>.

"inherently" employed a separate grammar, model, and lexicon. The Applicants respectfully disagree with these statements for the reasons stated below.

The Examiner's assertion that the Geilhufe system partitions a received command and analyzes the portions according to different grammars is contradicted by the express teachings of Geilhufe. Specifically, according to Geilhufe, ***an entire phrase*** is analyzed according to a ***single syntax***. See Geilhufe, col. 19, lines 15-17. Additionally, Geilhufe states that a ***single*** grammar—the standard VUI grammar-- is used to analyze all commands. Geilhufe, col. 8, lines 44-56.[2] Consequently, in the Geilhufe system, the same analysis (made according to the ***same*** syntax and the ***same*** grammar) is performed on the entire received command.

However, even if the Geilhufe system were to somehow split a command into multiple portions and analyze each portion separately, there is no "inherent" reason why the Geilhufe system would analyze each portion according to a ***different*** grammar, lexicon, and/or model as recited in claim 17. To take one example, since a lexicon is typically a dictionary of words and their pronunciation entries, a single lexicon could be used in systems where the commands are related to similar device types. And, in another example, a single lexicon might be used when the conservation of memory space was either required or advantageous.

For these reasons, the Applicants assert that Geilhufe fails to teach or suggest the switching between and use of multiple grammars, lexica and models as recited in claim 17. Consequently, for this additional reason, the Applicants assert that claim 17 is not anticipated by Geilhufe.

(B) Claims 1-16, 26-30, and 32-44 are not unpatentable over Junqua in view of Giuliani and White

As mentioned, claim 1 recites an interface control system for operating a plurality of devices. The system includes a 3 dimensional microphone array and a feature extraction module coupled to the microphone array. A speech recognition module is coupled to the

[2] Moreover, Geilhufe is completely silent as to the use of ***multiple*** grammars, lexica, or models.

feature extraction module and the speech recognition module utilizes hidden Markov models and can switch between different acoustic models and different grammars. At least one of the different acoustic models and at least one of the different grammars is downloaded over a network. A natural language interface module is coupled to the speech recognition module. A device interface is coupled to the natural language interface module and the natural language interface module operates a plurality of devices of one or more types that are coupled to the device interface based upon non-prompted, open-ended natural language requests from a user. The natural language interface module abstracts each of the plurality of devices into a respective one of the different grammars and a respective one of a plurality of lexica corresponding to each of the plurality of devices.

The Examiner stated that Junqua teaches all of the elements of claim 1 "but lacks explicitly wherein the natural language interface module abstracts each of the plurality of devices into a respective one of the different grammars and a respective one of a plurality of lexica corresponding to each of the plurality of devices." However, the Examiner stated that

> Geilhufe teaches an interface that abstracts...each of the plurality of devices (C.17.lines 6-10, C.19.lines 33-37, C.18.lines 1-4-wherein each device has "abstracted", core commands, and commands specific to a given application... it would have been obvious to modify Junqua's natural language parser and unified access controller with Geilhufe's device specific grammar and lexicon (vocabulary /specific list of commands). The motivation for doing so would have been to each device respond to specific commands appropriately (C.18.lines 1-4, 47-57- wherein "Aardvark call mom" results in calling mom from a desktop phone, by a command definition of a call as a specific command to a phone device, and, not, for example, a transcription of "Aardvark call mom" into a document.

In other words, the Examiner apparently asserts that Geilhufe teaches the abstracting of the devices into different grammars and lexica and that this alleged teaching can be used to modify the Junqua system. The Applicants respectfully disagree with these assertions for the reasons stated below.

In fact, Geilhufe does not teach the use of different grammars and lexica as is recited in claim 1. To the contrary, Geilhufe teaches that ***only a single grammar***—the standard VUI grammar -- is ever used. Geilhufe, col. 8, lines 44-56. Geilhufe is silent as to the use of either a single lexicon or multiple lexica.

Moreover, Geilhufe does not teach or suggest the use "device specific" grammars as asserted by the Examiner. Specifically, since the Geilhufe system uses a single grammar (i.e., the VUI grammar), this single grammar must be associated with ***all*** device types.

Consequently, since Geilhufe fails to teach or suggest the use of multiple grammars and/or lexica for any purpose and does not teach or suggest the use of device-specific grammars, the Applicants assert that claim 1 is allowable over the proposed combination.

Furthermore, even assuming Geilhufe somehow disclosed the use of different grammars and lexica, there is no suggestion or motivation to modify Junqua to have a natural language interface module that abstracts each of the plurality of devices into a respective one of the different grammars and a respective one of a plurality of lexica corresponding to each of the plurality of devices. There must be a motivation to make the proposed modification either in the references themselves or apparent to one skilled in the art. See MPEP § 2143.01.

More specifically, Junqua teaches a system that receives spoken instructions from a user. See FIG. 1 of Junqua, reproduced below for the convenience of the reader. The user's spoken instructions are converted into text by speech recognizer 20 and the output of the speech recognizer 20 is supplied to the natural language parser 26. Junqua, col. 2, lines 52-61. A natural language parser 26 then parses the text. *Id.* The output of the parser 26 is sent to a unified access controller 30, which sends electrical signals to activate a tuner 40 and a recorder 44.

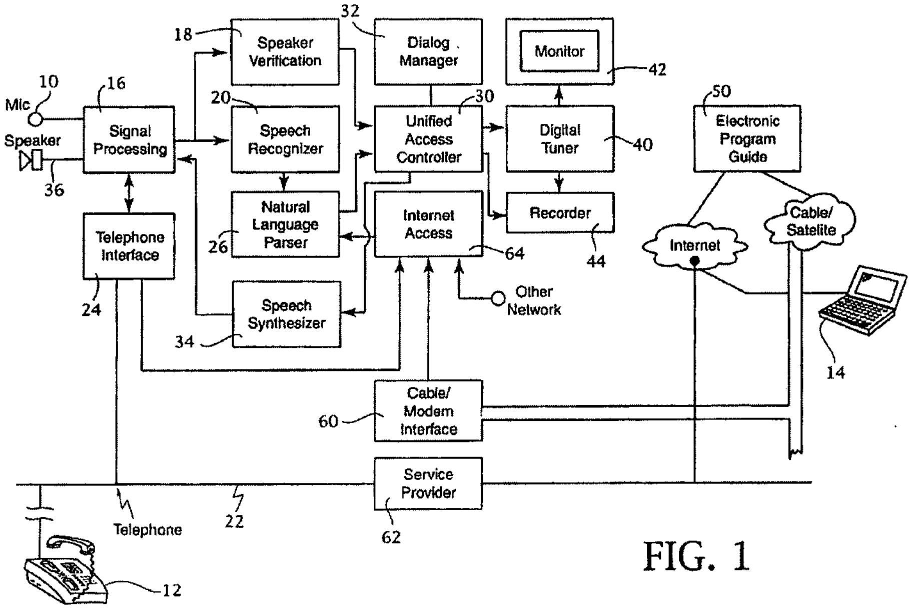

FIG. 1

The devices 40 and 44 controlled by the unified access controller 30 in the Junqua system are both of the same or similar type (i.e., relating to video technology). In fact, as can be seen in FIG. 1 of Junqua, other devices of differing types (i.e., the telephone 12 and computer 14) are not controlled by the unified access controller 30. Consequently, to achieve uniformity and design efficiency, it is submitted that one skilled in the art would be motivated to abstract each of the devices 40 and 44 according to the ***same*** grammar and lexicon. For instance, since the spoken words used to control each of the devices 40 and 44 are likely to be the same or similar (both relating to video technology), the same lexicon might advantageously be used. Simply put, there is no motivation either in the references themselves or apparent to one skilled in the art to modify the parser module 26 of Junqua in order to abstract each of the devices 40 and 44 into a different grammar and/or lexicon. To the contrary, if any motivation existed at all, it would be to abstract the devices 40 and 44 according to the same grammar and lexicon.

Consequently, for these additional reasons, the Applicants assert that the proposed modification of Junqua is non-obvious and assert that claim 1 is allowable over the proposed combination.

Claims 6, 7, 8, 9, 10, and 26 are independent claims that recite the use of open-ended requests and the switching between different grammars, lexicon, and models. Consequently, the Applicants assert that claims 6, 7, 8, 9, 10, and 26 are allowable for the same reasons as described above with respect to claim 1.

Claims 2-5, 11-15, 27-30, and 32-44 ultimately depend upon claims 1, 6, 7, 8, 9, 10, and 26, which have been shown to be allowable above, and therefore, these claims are also allowable. In addition, they introduce additional content that, particularly when considered in context with the claim from which they depend, introduce additional incremental patentable subject matter. Accordingly, the Applicants reserve the right to present further arguments in the future with regard to these dependent claims if independent claims 1, 6, 7, 8, 9, 10, and 26 are found to be unpatentable.